

ние самых современных мультимедиа технологий отображения информации в учебных материалах еще не гарантируют высокого уровня сформированности знаний и навыков по дисциплине. Не нужно забывать, что мультимедийный учебник или лабораторный практикум это всего лишь средство обучения, часть системы обучения по дисциплине.

На наш взгляд, первым шагом в направлении создания мультимедийных средств обучения должны стать разработка и создание преподавателем самой СИСТЕМЫ обучения по читаемой им дисциплине. В рамках разработки этой системы необходимо выделить организационный, содержательный, технологический и методический аспекты и сформировать модели, соответственно, организационного, содержательного, технологического и методического обеспечения системы. Тщательный анализ особенностей самой дисциплины, целей и задач ее изучения, выбор преподавателем уровней и критериев оценки качества ее усвоения позволяют грамотно определить роль и место мультимедийных средств в общей системе обучения по дисциплине, цели и средства их создания, виды и формы организации, программные средства реализации.

Якурнов А.С., Трофимов С. П.

СИСТЕМЫ ПОИСКА ИНФОРМАЦИИ НА ОБРАЗОВАТЕЛЬНЫХ ВЕБ-РЕСУРСАХ

alex@refo.ru

УГТУ-УПИ

г. Екатеринбург

Задачи, которые необходимо решить для реализации системы поиска обучающей информации:

- 1) определение факторов, по которым производится поиск информации на образовательных ресурсах;
- 2) выбор начальной точки для индексирования страниц, т.е. выбор начального тематического образовательного ресурса по интересующей информации или области знаний.

Для разработки собственной модели поисковой машины введем вектор параметров

$$\vec{S} = (p1, p2, p3, p4, p5, p6, p7)$$

$p1$ - число совпадений слов запроса с содержанием страницы;

$p2$ - число совпадений слов запроса с названием страницы;

$p3$ - число совпадений слов запроса с ключевыми словами страницы;

$p4$ - число совпадений слов запроса с описанием страницы;

$p5$ - число совпадений слов запроса с описанием картинок на странице;

$p6$ - индекс цитирования ресурса, которому принадлежит страница;

$p7$ - количество ссылок на данную страницу, совпадающих с запросом.

Одной из сложнейших задач для реализации поисковой машины является построение отображения набора параметров в скаляр

$$(p1, p2, p3, p4, p5, p6, p7) \rightarrow W ,$$

Пусть длина запроса равна N значимых слов.

Тогда основным параметром поиска является $p1$ (т.е. число совпадений слов запроса с содержанием страницы). Действительно, документ может рассматриваться как потенциально релевантный для данного запроса, если в его тексте найдены слова из поискового запроса.

Тогда первичный вид отображения будет выглядеть так:

$$W(p1, p2, p3, p4, p5, p6, p7) = \frac{p1}{N}$$

Это отображение уже является рабочим и часто используется в различных системах документарного поиска.

Следующими параметрами, которые позволяют улучшить качество поиска, являются параметры p_2, p_3, p_4 . Эти данные не являются самостоятельными, а призваны улучшить результаты поиска.

Рассмотрим эти параметры подробнее.

Параметр p_2 (число совпадений слов запроса с названием страницы) является самым важным, т.к. название страницы открыто пользователю и обычно отображается в верхней части веб-браузера. Более того, название страницы отображает посетителю тематику данного документа, его краткое содержание.

Поскольку лучше стремиться к нормализованному определению коэффициентов, т.е.

$W(p_1, p_2, p_3, p_4, p_5, p_6, p_7) \rightarrow 1$, а любой документ, для которого $W(p_1, p_2, p_3, p_4, p_5, p_6, p_7) = 1$, является абсолютно релевантным, то как оптимальное выбирается мультипликативное улучшение поиска.

Из-за отсутствия данных о посещаемости единственным критерием оценки популярности является индекс цитирования ресурса.

Параметры p_1, p_2, p_3, p_4 линейно возрастают в пределах $[0..N]$.

Индекс цитирования имеет степенной характер возрастания, т.е. в базе присутствуют ресурсы как ИЦ=1, так и ИЦ=20000.

Здесь нам также необходимо привести p_6 к $[c_4, 1]$.

Например, можно использовать такой вид:

$$w_6(p_6) = 1 + \lg\left(\frac{p_6}{MAX_IC}\right) / (\lg(MAX_IC))^K$$

Здесь MAX_IC - максимальный ИЦ в системе.

Тогда конечный вид отображения:

$$W(p1, p2, p3, p4, p5, p6, p7) =$$

$$\frac{p1}{N} \cdot (c1 + (1 - c1) \sqrt[n1]{\frac{p2}{N}}) \cdot (c2 + (1 - c2) \sqrt[n2]{\frac{p3^R + p4^R}{2N^R}}) \cdot (1 + \lg(\frac{p6}{MAX_IC}) / (\lg(MAX_IC))^K)$$

$c1, c2, R, n1, n2, K$ - константные параметры, которые позволяют настраивать модель поиска под конкретные технические требования поисковой системы.

Выбранные коэффициенты позволяют начать исследовать результаты поиска и нуждаются в корректировке на основе получаемых данных в ходе тестирования системы.

Для индексации образовательных ресурсов:

- - в качестве начального зерна для индексации страниц выбирается рубрика “Образование” Яндекс каталога или уже имеющийся список ссылок на тематические образовательные интернет ресурсы;
- - устанавливается глубина индексирования внешних ресурсов равная 1, т.е. поисковая система индексирует все образовательные ресурсы в Яндекс каталоге и ссылки с этих ресурсов, но не далее;
- - для улучшения качества отбираемых системой ресурсов и более точного соответствия информационных ресурсов теме дальнейшего поиска, при входе на новые ресурс используется фильтр по ключевым словам в названии или тексте документа.